NETWORK INTRUSION DETECTION WITH NAÏVE BAYES CLASSIFICATION AND SELF ORGANIZING MAPS

Master's Student: Mubeen Iqbal

Supervised by: A/Prof. Quang Ha

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY (FEIT)

UNIVERSITY OF TECHNOLOGY SYDNEY (UTS)

August 2014

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Mubeen Iqbal

Acknowledgment

First and foremost, I offer my sincerest gratitude to my advisor, **A/Prof. Quang Ha**, who has backed me, all around my thesis with his perseverance and knowledge. This research would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

Secondly, I would like to thank my research colleagues for their encouragement throughout my research work.

Finally, I would like to thank my parents for their unending love, support and understanding during my research work.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

IDS	Intrusion Detection System
LSHSN	Large-Scale High-Speed Networks
KDD	Knowledge Discovery in Data Competitions
ITS	Intrusion Tolerant System
NIDS	Network Intrusion Detection System
MADAM	Mining Audit Data for Automated Model
HIDS	Host Based Intrusion Detection System
IDES	Intrusion Detection Expert System
NIDES	Network Intrusion Detection Expert System
SVM	Support Vector Machines
SOM	Self Organizing Map
NB	Naïve Bayes
ML	Machine Learning
DARPA	Defense Advanced Research Projects Agency
HMM	Hidden Markov Model
HTML	Hyper Text Markup Language
DOS	Denial-of- Service
U to R	User to Root Attacks
R to L	Remote to Local Attacks
DR	Detection Rate
FPR	False Positive Rate
WEKA	Waikato Environment for Knowledge Analysis

Abstract

In this digital period, internet has turned into an indispensable wellspring of correspondence in just about every calling. With the expanded use of system engineering, its security has developed to be exceptionally discriminating issue as the workstations in distinctive association hold very private data and touchy information. The system used to screen the system security is known as Network detection. Intrusion detection is to get ambushes against a machine structure. It is a discriminating enhancement great to go part and additionally an element extent of examination. In Information Security, Intrusion recognizable proof is the showing of placing exercises that attempt to deal the protection, respectability or availability of a benefit. It accepts an astoundingly key part in waylay area, security check and framework inspect. One of the vital tests to Intrusion Detection is the issue of misjudgement, misdetection and unsuccessful deficiency of steady response to the strike. In the past years, as the second line of boundary after firewall, the Intrusion Detection strategy has got speedy progression.

This research work prepares two diverse Machine Learning techniques, both supervised and unsupervised, for Network Intrusion Detection. These techniques are Naïve Bayes (supervised learning) and Self Organizing Maps (unsupervised learning). The KDD Cup 99 dataset is utilized for Intrusion Detection Problem. As KDD Cup 99 dataset holds some symbolic attribute and also numeric attributes, two sorts of transformation technique have been utilized for these properties. These are conditional probabilities conversion technique and indicator variables transformation. The two machine learning procedures are prepared on both kind of transformed dataset and afterward their outcomes are looked at with respect to the correctness of intrusion detection.

Keywords: Network Intrusion Detection, supervised learning, unsupervised learning, Self-Organizing Map, Naïve Bayes, Conditional Probability Symbolic Conversion, Indicator Variable Symbolic Conversion, KDD Cup 1999.